DOCUMENT RESUME

ED 446 150 TM 031 916

AUTHOR Kenney, Patricia Ann; Silver, Edward A.
TITLE Design Features for the Content Analysis of a State Assessment and NAEP.
INSTITUTION Pittsburgh Univ., PA. Learning Research and Development Center.
SPONS AGENCY National Assessment Governing Board, Washington, DC.
PUB DATE 1999-00-00
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999). For a related paper about the content analysis study, see TM 031 915.
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *Content Analysis; *Grade 8; Junior High Schools; *Mathematics Tests; National Competency Tests; *Research Design; Test Construction; Test Content; Test Items; *Test Results
IDENTIFIERS Experts; *Maryland School Performance Assessment Program; *National Assessment of Educational Progress

ABSTRACT

This paper presents an overview of the design features that were developed for the Content Analysis Project. The purpose of the project was to examine the congruence between a state's test in eighth-grade mathematics and that used by the National Assessment of Educational Progress. The results of this analysis were then to be used to determine whether the identified differences between the state assessment and the NAEP were sufficient to account for the magnitude of the differences between proficient performance on the two tests. A second purpose was to develop a model process states could use to compare their frameworks and assessments to NAEP in mathematics and in other content areas, such as reading and science. The paper presents details about three design features of the comparison study: (1) the use of expert judgment; (2) the importance of viewing the test from multiple perspectives; and (3) the implementation of a multiphase process for comparing the tests. Some of the limitations of the study are outlined. An appendix summarizes the phases of project development. (Contains 20 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 446 150

Design Features for the Content Analysis of a State Assessment and NAEP

Patricia Ann Kenney and Edward A. Silver

Learning Research and Development Center
University of Pittsburgh

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Preparation of this paper was supported by a contract to the University of Pittsburgh's Learning Research and Development Center from the National Assessment Governing Board. Any opinions expressed herein are those of the authors and do not necessarily reflect the views of the Board. This paper was presented at the 1998 Annual Conference of the American Educational Research Association, Montreal, Quebec, Canada.

TM031916

BEST COPY AVAILABLE

Design Features for the Content Analysis of a State Assessment and NAEP

Patricia Ann Kenney and Edward A. Silver[1]

Learning Research and Development Center
University of Pittsburgh

In this paper, we present an overview of the design features that were developed for the Content Analysis Project, a study funded by the National Assessment Governing Board (NAGB). The purpose of this project was two-fold: (1) to examine the congruence[2] between a state's test in eighth-grade mathematics and that used by NAEP and then use the results to answer a fundamental question: Are identified differences between the state assessment and NAEP sufficient to account for the magnitude of difference between proficient performance on the two tests? and (2) to develop a model process that states could use to compare their frameworks and assessments to NAEP not only in mathematics but also in other content areas (e.g., reading, science). The paper consists of three sections, with the first section focusing on background information about the Content Analysis Project. Next, we present details about three design features of the study that compared NAEP and the state test: the use of expert judgment, the importance of viewing the test from multiple perspectives, and the implementation of a multi-phase process for comparing the tests. We conclude with the limitations of the study and some comments.

Background

[1]We would like to acknowledge the valuable contributions of two colleagues to the work described in this paper: Dr. Judith S. Zawojewski (Associate Professor of Mathematics Education, National-Louis University) and Dr. Cengiz Alacaci (Post-doctoral Research Associate, LRDC).

[2]In mathematics, two geometric figures are said to be congruent if they can be superimposed so as to coincide, and there are a number of ways that congruence can be demonstrated mathematically. Here, we use the word "congruence" as a synonym for the relationships between important components of each test (e.g., the congruence between the state framework and the NAEP framework). Absolute judgments about congruence were not possible, but we believed that it would be possible to describe the congruence (or lack of it) between the two tests.

In its Redesign Policy (1996), NAGB outlined a number of goals and objectives for guiding changes in the National Assessment of Educational Progress (NAEP). One particular set of goals and objectives involved assisting states in linking their assessments with NAEP, and several states have already begun to establish such links. In particular, some states have begun to report student performance in terms of state-level "proficiency" standards, employing language similar to that used in the NAEP achievement levels. For many states that participate in NAEP, however, discrepancies have emerged between the percentages of students scoring at the NAEP proficient level and those meeting the state standard for proficient performance (Musick, 1996; Archer, 1997). In general, the trend is that the percentage of students meeting the state standard for proficient performance as defined by the state is higher than that of students in the state NAEP sample who meet the "proficient" achievement level as defined by NAEP.

What factors contribute to these differences in proficient performance on a state's assessment and proficient performance on NAEP? There are many possible reasons for the performance differences including variations in the purposes of the assessments, in the definitions of "proficient" and the processes used to set proficiency standards, or in content coverage between the state test and NAEP. Musick (1996) proposed that it is important to examine the state assessment programs and NAEP in order to identify the possible reasons for these differences.

Based on Musick's report and on conversations with state policy makers, NAGB funded a study that would address possible reasons for the differences in performance levels. The focus of the NAGB-sponsored study was on the tests[3] themselves and did not involve issues about how the proficiency levels were defined and set. As noted previously in this paper, the study examined the congruence between a state's test in eighth-grade mathematics and that used by NAEP and then used the results to answer a fundamental question: Are identified differences between the state assessment and NAEP sufficient to account for the magnitude of difference

[3]In this report, the terms "test" and "assessment" are often used interchangeably, following Shepard (1994). If there is a difference between these two terms, it is one of emphasis: a test usually refers to a particular coherent test instrument; an assessment is more likely to refer to a system that involves more than one test.

between proficient performance on the two tests? Additionally, an important part of the study was the development of a model process that states could use to compare their frameworks and assessments to NAEP not only in mathematics but also in other content areas (e.g., reading, science).

Two states, North Carolina and Maryland, participated in the study. For both states, there was a discrepancy between performance on the state test and NAEP (as reported in Musick, 1996):

- For the Maryland School Performance Assessment Program (MSPAP) test in eighth-grade mathematics (1994-95), 48% of the students met the state proficiency standard as compared to 24% of the students in the state-NAEP sample in 1996 performing at the proficient achievement level.

- For the North Carolina End-of-Grade test in mathematics at grade 8 (1994-95), 68% of the students met the state proficiency standard as compared to 20% of the students in the state-NAEP sample in 1996 performing at the proficient achievement level.

Among the reasons for including these two particular states in the study were that representatives from each state expressed strong interest in the project, and that the assessments used in each state are quite different in a number of important ways. One important way in which the assessments differed was format: the North Carolina assessment is composed entirely of multiple-choice questions and the Maryland assessment is composed entirely of constructed-response questions. It was thought that the diversity in format as well as other ways in which the state assessments differed from each other (e.g., purpose; reporting level) would contribute to the generalizability of the content analysis process.

Design Features of the Content Analysis Project

In thinking about the design of a process to investigate the congruence between a state's test and NAEP, we considered three features to be especially important: 1) the process should involve consensus judgments by a panel whose members were selected on the basis of their

expertise in areas relevant to the study (e.g., middle school mathematics, the state test, NAEP); 2) the state test and NAEP should be examined from multiple perspectives (e.g., technical characteristics; content areas; cognitive demand) and according to an array of aspects (e.g., test frameworks and specifications; test items; scoring guides and student work); and 3) the process should be multi-phased (i.e., there should be adequate time for us to prepare materials and analyze data and for the panelists to discuss important issues and to reach consensus). Each of these design features is discussed next.

The Consensus Judgments of a Panel of Experts

In recent years, basing decisions about NAEP on expert judgment has become a common occurrence. For example, expert judgment about student performance is at the heart of the achievement levels-setting process (NAGB, 1990). For the NAEP mathematics assessment, judgments of mathematics education professionals were used to establish the content and curricular validity of the tests that comprised the trial state assessments in 1990 and 1992 (Silver & Kenney, 1994; Silver, Kenney, & Salmon-Cox, 1992) and to examine the 1992 NAEP achievement levels-setting efforts (Silver & Kenney, 1993). For this study, we used a panel of experts to assist in making the congruence judgments for each state assessment and NAEP.

The panel of experts charged with examining the relationship between a state's assessment and NAEP was composed of six mathematics education professionals (e.g., mathematics teachers, college/university mathematics educators, mathematics curriculum specialists), and the composition of the panel reflected distributed expertise that spanned the state test, NAEP, and middle school mathematics. Of the six members, two members were selected on the basis of their familiarity with the state assessment; that is, they served in a capacity that ensured knowledge of the state's testing program (e.g., serving on the mathematics framework development committee; writing test items; providing professional development for mathematics teachers on the state assessment program). Personnel from the state's department of education nominated possible panelists, and we contacted them. Having representatives from the state as members of the panel ensured that states were an integral part of the content analysis process. --

Also, the two "state" panelists served as resource people when informational questions arose about the state test.

Another pair of panelists was selected on the basis of their knowledge of the NAEP mathematics assessment, and in particular the NAEP grade-8 test. For example, these panelists had served on committees that developed the NAEP mathematics framework and items, or they knew about NAEP through their involvement with other NAEP-related projects such as the NCTM NAEP Interpretive Reports Project (Kenney & Silver, 1997; Silver & Kenney, in press). The two "NAEP" panelists could provide the group with expertise about that test, should the need arise.

The last two panelists were selected for their expertise about and experience with middle school mathematics education and for their lack of specialized knowledge about either the state assessment or about NAEP. The role of these panelists within the group was one of neutrality with respect to the tests to be examined; that is, this pair of "neutral" panelists had no vested interest in either test.

The six panelists met for two days in two separate sets of meetings held about a month apart. The structure of the panel -- two state panelists, two NAEP panelists, two neutral panelists -- allowed varying points of view to emerge during the discussions. Also, in instances where the panelists would work in small groups, it was possible to form two subgroups of three members, (one NAEP, one state, one neutral). It is important to note here that different six-member panels were selected for each state-NAEP comparison; that is, no panelists served on both the North Carolina and Maryland panels. This was done to ensure that direct comparisons would not be made between the state tests, but only about the state test and NAEP.

In addition to the six members of the panel, there were others who participated in the consensus process and who brought with them additional expertise to the consensus process. For example, state testing directors, testing consultants, and mathematics specialists from the states were invited to participate in the activities and discussions at the meetings. Members of the NAGB staff and a member of NAGB (Mark Musick) also were involved in some of the

deliberations. Representatives from the National Center for Education Statistics (NCES) also attended some meetings. Finally, in addition to providing additional expertise about NAEP, we were responsible for creating all materials used at the meetings, analyzing data, and serving as facilitators of the panelists' and other participants' discussions about the congruence between the tests.

The Tests as Viewed from Multiple Perspectives

In formulating the design for this study, we proposed that tests could be compared according to multiple perspectives, hereafter referred to as "dimensions." Three dimensions common to the state test and NAEP were identified as relevant to this study. First, there is a technical dimension that involves components such as the number and type of items, the time allotted to administering the test, the difficulty of the items, etc. A second aspect involves a content dimension that has to do with the particular content topics (e.g., for mathematics -- geometry, measurement, algebra) included. And a cognitive dimension involves the extent to which a test engages students in various cognitive processes, including problem solving, reasoning, or the recall of facts and definitions. Because each test has a distinct profile with respect to these dimensions, it is possible to determine the profile for each test and then to compare the tests for congruence on all three dimensions. A model for this process appears in Figure 1. The methods used to examine the technical, content, and cognitive aspects of each test are described next.

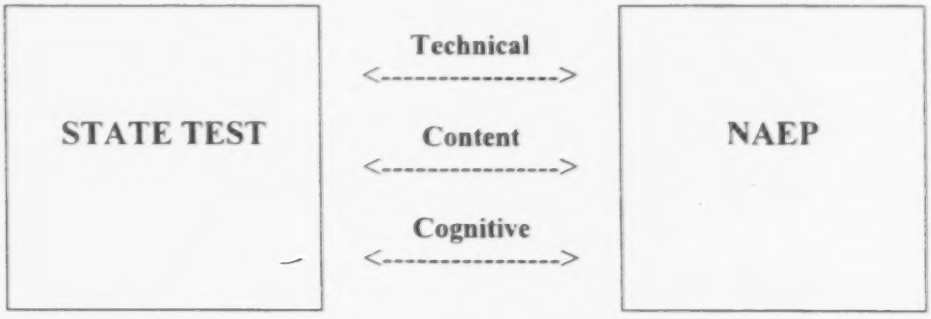

Figure 1 Dimensions along which to investigate congruence between the state test

Technical dimension. The technical aspect involves particulars such as the number of items on the test, the type of items (multiple choice; constructed response), the time allotted for students to take the test, the difficulty of the items, etc. Information about a test's technical characteristics most often appears in documents such as frameworks and specifications and the testing program's technical reports. It was deemed important to this study that the panelists and other participants be knowledgeable about the technical aspects of the state assessment and NAEP before beginning their comparison of the tests themselves.

In this study, information about the technical aspects of the state assessment and that of NAEP was obtained in two ways. First, using the framework documents and technical reports, we prepared summaries of technical information about each test to be used at the panel meetings and in the project's final report. These summaries were verified for accuracy by the representative from the state's department of education and by members of the NAGB staff.

The second way that technical information was obtained was through presentations made at the first meeting. In particular, a representative from the state's department of education presented an overview of the purpose of the state test and its important technical characteristics; the first author of this paper, who is very familiar with NAEP, gave a similar presentation about that assessment, with the NAGB staff members providing additional information when necessary. These presentations enabled the panelists and other participants to get further clarification on the technical aspects of each test.

Content dimension. The content aspect of a test has to do with what is being assessed; that is, in mathematics this involves the particular topics included on the test (e.g., number concepts and relationships; measurement; geometry; statistics and probability; algebra). These content topics are common to most mathematics assessments, but the content coverage across topics can vary widely depending on the purpose of the test. For example, on a basic competency test a large percentage of the items could be devoted to topics in number properties

and computation, with lower percentages of items in the other content areas. Another grade-8 test, which has a more broadly defined purpose, could include items that are equally distributed across content areas. Thus, for the tests just described, although both include the same content topics, the coverage is different, thus likely affecting the content congruence between the tests.

The content aspects of a state test and NAEP and the congruence between them were investigated in two ways: a framework-to-framework matching by content area and an item-to-framework cross-matching of items from one test onto the framework of the other test. The framework-to-framework matching activity involved comparing carefully the content topics as presented in the 1996 NAEP mathematics framework document (College Board, 1994) and those presented in the relevant mathematics "framework" document from the state. For a state, the relevant framework document is most likely the state's curricular goals for mathematics at each grade level or clusters of grade levels (e.g., grades 6-8), and there is evidence that many states use the curricular goals as the test specifications for their testing programs (Roeber, Bond, & Braskamp, 1997). Because both the NAEP mathematics framework and the state curricular goals are based in large part on content topics, using these documents in the framework-to-framework activity was deemed reasonable. The activity itself involved the panelists and other participants identifying topics in both the NAEP framework and the state framework that were similar, topics that were in NAEP but missing in the state framework, and topics that were in the state framework but missing in NAEP. Comparing the common topics in each framework and the topics unique to each framework provided a way to evaluate the congruence between NAEP and the state test on the basis of intended content coverage, as specified in the frameworks.

The item-to-framework, cross-matching activity involved having the panelists and other participants classify NAEP items according to the content topics in a state's framework document and items from the state test according to the NAEP framework. This activity was designed to serve two purposes. First, it provided an additional opportunity to examine the items from each test. Additionally, the results from the activity can be used to validate the information from the framework-to-framework matching through "triangulation" of the data. In qualitative research,

triangulation is a standard technique that draws on multiple methods and data sources to gain more confidence in the accuracy of the findings (Jick, 1983; Mathison, 1988). For example, in the context of this study, suppose that the framework-to-framework matching activity revealed that the measurement topic of converting units within the same system (e.g., inches to feet; millimeters to centimeters) appears in both the NAEP mathematics framework and in the state's framework document. Then, it should be highly likely that in the item-to-framework cross-matching activity, NAEP items classified as assessing conversion of units should be classified in the state framework's category associated with conversion of units, and vice versa. If the expected classification occurred, then the framework-to-framework match was confirmed. However, if the item classification went in another direction (e.g., the item was classified in an unexpected category), then the outcome would be "non-confirmation" of the framework-to-framework match, and reasons for this non-confirmation could be explored

Results from the two activities just described allowed the panelists and other participants to evaluate the congruence between the NAEP and state tests with respect to their content dimension. Additional information about the congruence came from discussions among the panelists and participants as they completed each activity. We along with two other colleagues from LRDC served as the facilitators of these discussions.

Cognitive dimension. The cognitive aspect of a test refers to the extent to which a test engages students in various cognitive processes such as recalling important facts and definitions, computing with numbers, demonstrating conceptual understanding, and using reasoning in mathematical situations. In designing this study, we recognized the importance of comparing the tests with respect to the cognitive demands each test placed on students, the premise being that even though two tests might be similar in terms of what content topics are included, they could be quite different on how the topics were assessed. How topics are assessed on the test goes well beyond content area and item format considerations and into the realm of whether the focus is on lower-order skills such as recall of facts and routine procedures or on higher-order skills such as problem solving and mathematical reasoning, or a combination of both kinds of skills.

The cognitive aspects of the NAEP and the state test were compared on the basis of two activities. First, we chose a set of criteria external to both assessments that could be used to evaluate the cognitive demand of the items on each test. The criteria were obtained from sources such as the Curriculum and Evaluation Standards for School Mathematics (NCTM, 1989) and other studies involving NAEP (e.g., Romberg, Smith, Smith, & Wilson, 1992; Silver & Kenney, 1994), and the criteria represented both high-level (problem solving, communication, reasoning) and low-level (recall of facts, routine procedures) cognitive processes. The panelists and other participants used these criteria to evaluate items from NAEP and the state assessment. The findings from this part of the investigation were based on the results of the evaluation of the items according to cognitive demand and on a discussion among the panelists and the others at the meeting.

Because the Maryland assessment program uses a test composed completely of constructed-response items, it was important for the comparison between that test and NAEP to focus on the cognitive demand of those items along with their scoring guides and sample student responses at each score level. In particular, the design of this activity was based on this idea: If the cognitive demand of the item is high, then is that high cognitive demand sustained in the scoring guide for that item and in the set of sample student responses for each score level? The panelists and other participants had the opportunity to examine carefully some constructed-response items from NAEP and from the Maryland test. The issues concerning the cognitive demand of the items and whether that demand was sustained in the scoring guides and student work were then discussed by the group.

The Multi-phase Process

Because the process involved expert judgment concerning the congruence of the state test and NAEP along multiple dimensions, it was important to plan carefully the sequence of events so that the panelists would have adequate time to examine each test completely, to discuss important issues as a group, and to reach consensus. Also, we needed time to analyze the data generated by the panelists, to synthesize the results of the group discussions, and to plan ways in

which to share information with the panelists so as to inform their judgments. Based on these considerations, it was decided that the process should be multi-phased, with five distinct phases: pre-meetings, first panel meeting, between meetings, second panel meeting, and post-meetings.

Appendix A contains an out of the five phases and a brief summary of the activities occurring in each phase. Two of the five phases (Phases II and IV) involved the activities occurring during the two-day meetings of the panel of experts. In general, considering the technical and content aspects of the state test and NAEP was the focus of the first meeting; examining the cognitive aspects of the tests was the focus of the second meeting, with the final portion of that meeting devoted to a discussion of the question concerning whether differences in the technical, content, or cognitive dimensions were sufficient to account for the performance differences at the proficient level between the state test and NAEP. The other three phases (Phases I, III, and V) allowed us to obtain and study relevant documents and other materials from NAEP and from the state assessment, to prepare materials for the meetings, to analyze data generated during the meetings, and to produce summaries for the panel meetings and the final report.

Limitations of the Study

As stated previously, the purpose of this study was to examine the congruence between a state's test in eighth-grade mathematics and that use by NAEP to answer a fundamental question: Are identified differences between the state assessment and NAEP sufficient to account for the magnitude of difference between proficient performance on the two tests? This specificity of purpose imposed these limitations on the study:

1) **There was no direct attempt to evaluate either the state assessment or NAEP as a part of this study.** Instead, we assumed that the state assessment was carefully developed and had undergone some kind of evaluation, and we looked for documentation (e.g., technical reports, research studies) that supported these assumptions. In the case of NAEP, there is evidence that it has been extensively evaluated by external groups such as the National Academy

of Education (1992, 1994) and more recently the National Academy of Sciences (National Research Council, 19989

2) The study stopped short of comparing the tests according to the ways in which results were reported. Our charge was to consider only the tests themselves according to technical, content, and cognitive characteristics and for constructed-response questions, the way in which such questions were scored. Comparing the tests according to reporting issues might be the focus of another study.

3) The study was concerned exclusively with the subject area of mathematics. Although we made a conscious attempt to use design principles that could be applied to subject areas other than mathematics (e.g., reading; science), the extent to which the design principles actually can be used to evaluate the congruence between tests in other subject areas should be established in future studies.

Concluding Comments

In this paper, we have summarized background information concerning the perceived need for this study and presented an explanation of and rationale for the key features of the study design. The design features of consensus judgment by a panel of experts, examination of the state test and NAEP along multiple dimensions, and the organization of the process into multiple phases were selected as not only being relevant to the study of the congruence between two mathematics tests, but also because it was thought that these features would generalize to other content areas and grade levels assessed by NAEP and by states. For example, with regard to the panel of experts, the members can be selected according to their expertise in other disciplines such as reading or science and according to their expertise at the elementary, middle school, or high school levels. And it is likely that any state test and the NAEP test in a discipline other than mathematics can be evaluated along the three dimensions -- technical, content, and cognitive -- described in this section, although the details would vary by discipline. The multi-phase process, which is discipline-independent, serves as a suggested structure for the study itself. These three design features, then, can contribute to the development of a model process for examining the

congruence between a state test and NAEP that can be used in disciplines other than mathematics and at grade levels other than grade 8. We also presented some important limitations to the study.

References

Archer, J. (1997, January 15). States struggle to ensure data make the grade. Education Week, pp. 1, 30.

College Board. (1994). Mathematics framework for the 1996 National Assessment of Educational Progress. Washington, D.C.: National Assessment Governing Board.

Jick, T. J. (1983). Mixing qualitative and quantitative methods: Triangulation in action. In J. V. Maanen (Ed.), Qualitative methodology. Beverly Hills, CA: Sage.

Kenney, P. A., & Silver, E. A. (Eds.) (1997). Results from the Sixth Mathematics Assessment of the National Assessment of Educational Progress. Reston, VA: National Council of Teachers of Mathematics.

Mathison, S. (1988). Why triangulate? Educational Researcher, 17(2), 13-17.

Musick, M. D. (1996, June). Setting educational standards high enough. Atlanta: Southern Regional Education Board.

National Academy of Education. (1992). Assessing student achievement in the states. Stanford, CA: The Academy.

National Academy of Education. (1994). The Trial State Assessment: Prospects and realities. Stanford, CA: The Academy.

National Research Council. (1999). Grading the Nation's Report Card: Evaluating NAEP and transforming the assessment of educational progress. Washington, D. C.: National Research Council.

National Assessment Governing Board. (1996, August). Policy statement on redesigning the National Assessment of Educational Progress. Washington, D.C.: Author.

National Assessment Governing Board. (1990, May). Setting appropriate achievement levels for the National Assessment of Educational Progress. Washington, D. C.: Author.

National Council of Teachers of Mathematics. (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: Author.

National Research Council. (1999). Grading the Nation's Report Card: Evaluating NAEP and transforming the assessment of educational progress. Washington, D. C.: National Academy Press.

Roeber, E., Bond, L., & Braskamp, D. (1997). Annual survey of state student assessment programs, Fall 1996. Washington, D.C.: Council of Chief State School Officers.

Romberg, T. A., Smith, M., Smith, S., & Wilson, L. (1992, June). The feasibility of using international data to set achievement levels for the National Assessment of Educational Progress (NAEP). Madison, WI: National Center for Research in Mathematics Education.

Shepard, L. (1994). The challenge of assessing young children. Phi Delta Kappan, 76(3), 206-210.

Silver, E. A., & Kenney, P. A. (1993). Expert panel review of the 1992 NAEP Mathematics Achievement Levels. In National Academy of Education, Setting performance standards for student achievement: Background studies (pp. 215-282). Stanford, CA: The Academy.

Silver, E. A., & Kenney, P. A. (1994). The content and curricular validity of the 1992 NAEP Trial State Assessment (TSA) in mathematics. In National Academy of Education, The Trial State Assessment: Prospects and realities: Background studies (pp. 231-284). Stanford, CA: The Academy.

Silver, E. A., & Kenney, P. A. (Eds.) (In press). Results from the Seventh Mathematics Assessment of the National Assessment of Educational Progress. Reston, VA: National Council of Teachers of Mathematics.

Silver, E. A., Kenney, P. A., & Salmon-Cox, L. (1992). The content and curricular validity of the 1990 NAEP Mathematics Items: A retrospective analysis. In National Academy of Education, Assessing student achievement in the states: Background studies (pp. 157-218). Stanford, CA: The Academy.

Appendix A

The Five Phases of the Content Analysis Project

Phase I: Before the First Meeting

- LRDC project staff (i.e., Kenney, Silver, Zawojewski, Alacaci) gathered information on the state test and NAEP and prepared handouts and focus questions to be sent to the panelists.
- LRDC project staff compiled information on the technical aspects of each test.
- LRDC project staff prepared activities for the first meeting concerning the tests' content aspects.
- Panelists read handouts and responded to the focus questions.

Phase II: First Meeting

Day 1

- Representative from the state and from NAEP gave presentations on their respective assessment programs.
- Panelists discussed their responses to the focus questions completed prior to the meeting; LRDC project staff served as facilitators for this discussion.
- Panelists worked individually and then in small groups on the framework-to-framework matching activity; LRDC project staff served as facilitators in the small group discussions.

Day 2

- Panelists discussed the findings from the framework-to-framework matching activity and reached consensus on the congruence between the two tests based on content characteristics.
- Panelists worked individually on the item-to-framework cross-matching activity (NAEP items to state framework; state items to NAEP framework)

Phase III: Between the Meetings

- LRDC project staff analyzed the results of the item-to-framework matching activity to validate the judgments of the panelists from the framework-to-framework matching activity.
- LRDC project staff prepared a summary of the content congruence decisions to share with the panelists.
- LRDC project staff compiled information about criteria that was used to evaluate the cognitive aspects of the tests and sent it to the panelists; project staff also prepared materials for use during the second meeting.
- Panelists received and read information about the criteria to be used at the meeting to evaluate the cognitive aspects of the tests.

Phase IV: Second Meeting

Day 1

- LRDC project staff shared findings from the activities completed at the last meeting and facilitated a discussion by the panelists on the content congruence between the state test and NAEP.
- Panelists worked individually on an activity that asked them to evaluate the cognitive demands of a set of NAEP items and a set of items from the state test.
- LRDC project staff produced a preliminary analysis of the data from the cognitive demand activity for presentation at Day 2 of the meeting.

Day 2

- LRDC project staff shared the preliminary findings from the cognitive demand activity with the panelists and facilitated a discussion of those findings.
- [for the Maryland-NAEP meetings]. Panelists completed an activity concerning the level of cognitive demand as sustained from constructed-response item to scoring guide to examples of student work at each score level. LRDC project staff facilitated the discussion based on this activity.
- Panelists engaged in a discussion, facilitated by the project staff, concerning the congruence between the tests on their cognitive characteristics.
- Based on their judgments about the congruence between the tests on the three dimensions (technical, content, cognitive), the panelists worked to reach consensus on the differences between the state test and NAEP and whether these differences were sufficient to account for the magnitude of difference between proficient performance. Panelists also suggested other factors that could be contributing to the performance differences.

Phase V: After the Meetings

- LRDC project staff prepared a report that summarized the findings from the project and submitted that report to the state and to NAGB.